2

AD-A214 241

215400-1-F

Final Report

SURVEY OF ERIM APPROACHES APPLICABLE TO SEMI-AUTOMATIC TARGET DETECTION AND CUEING FOR MULTISPECTRAL AND MULTISENSOR EXPLOITATION

L. WOOD

JULY 1989

PAR Government Systems Corp.
220 Seneca Turnpike
New Hartford, NY 13413

ERIM
P.O. Box 8618
Ann Arbor, MI 48107-8618

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. 215400-1-F	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Survey of ERIM Approaches Applicable to Semi-automatic Target Detection and Cueing for Multispectral and Multisensor Exploitation		5. Report Date 25 July 1989
		6. Performing Organization Code
7. Author(s) Lynnette Wood		8. Performing Organization Report No.
9. Performing Organization Name and Address ERIM P.O. Box 8618 Ann Arbor, MI 48107-8618		10. Work Unit No.
		11. Contract or Grant No. SC-88-0151-02
12. Sponsoring Agency Name and Address PAR Government Systems Corporation 220 Seneca Turnpike New Hartford, NY 13413		13. Type of Report and Period Covered Final Report
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Tactical and strategic decisions must increasingly be made based on the aggregation and integration of information from various sources. Multiple sensors can be employed to provide a range of parameters which can aid in identifying enemy targets. The synergistic combination of data from these various sensors, as well as from other sources, can enhance a photointerpreter's ability to locate and identify targets. A semi-automatic multispectral multisensor system for target detection would be invaluable for this purpose. The Semi-Automated Multispectral/Multisensor Exploitation (SAMME) Project is an effort to develop computer assisted algorithms and display methods to aid interpreters in identifying targets in multispectral multisensor data.

This report outlines approaches attempted at ERIM which are relevant to the SAMME project. A series of over 200 ERIM documents were reviewed. Included in the review were internal memos, technical reports, white papers, proposals and journal articles published by ERIM employees which describe results of work done at ERIM. The report includes information of benefits and drawbacks of various approaches under different circumstances, when this information is available. However, this report is not a recommendation for a particular system, only a compendium of knowledge applicable to the task.

17. Key Words Image data fusion Multispectral Multisensor Target detection	18. Distribution Statement This document has been approved for public release and sale; its distribution is unlimited.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages	22. Price

215400-1-F

Final Report

SURVEY OF ERIM APPROACHES APPLICABLE TO SEMI-AUTOMATIC TARGET DETECTION AND CUEING FOR MULTISPECTRAL AND MULTISENSOR EXPLOITATION

L. WOOD

JULY 1989

PAR Government Systems Corp.
220 Seneca Turnpike
New Hartford, NY 13413

TABLE OF CONTENTS

TABLE OF CONTENTS (Concluded)

LIST OF FIGURES

1.0 INTRODUCTION

Tactical and strategic decisions must increasingly be made based on the aggregation and integration of information from various sources. Multiple sensors can be employed to provide a range of parameters which can aid in identifying enemy targets. The synergistic combination of data from these various sensors, as well as from other sources, can enhance a photointerpreter's ability to locate and identify targets. A multispectral multisensor system for semi-automated target detection would be invaluable for this purpose.

Two goals of such a system are to increase the probability of detecting and locating targets of interest and to reduce false alarms. By combining the data from several sources, each having their own unique strengths and weaknesses, one hopes to "fuse" the data in a way to better discriminate targets. The data itself may consist of imagery from several different sensors, including infrared, radar and visible light, as well as cartographic and intelligence information. The data from these various sources should be combined a way to make the information easily understood and optimally exploitable.

One benefit of this type of approach is that multisensor multidiscriminant systems will frustrate countermeasure efforts (camouflage, concealment and deception). For instance, a target hidden from radar through the use of scattering nets may still be detectable with thermal infrared detectors. Also, collection of data can continue regardless of time of day or weather conditions; infrared and SAR data can be collected at night and SAR data can be collected in poor weather. Finally, multisensor data can be used to reduce the volume and complexity of information to be interpreted.

Under Contract Number F30602-88-C-0151, PAR Government Systems Corporation (PGSC) is performing the Semi-Automated Multispectral Multisensor Exploitation (SAMME) program for Rome Air Development Center (RADC). The objective of the SAMME program is to demonstrate semi-automated aids for the image interpreter to more fully exploit imagery by using multi-source information to detect and identify targets. Under subcontract to PGSC, ERIM is providing support in requirements analysis, system design, and demonstration and test. ERIM system design support consists of a study of appropriate techniques developed or validated at ERIM and a recommendation of specific approaches to use in the SAMME program.

This report outlines approaches attempted at ERIM which are relevant to computer assisted algorithms and display methods which aid photointerpreters in using multispectral multisensor imagery in conjunction with multisource data to detect and identify targets. A series of over 200 ERIM documents were reviewed. Included in the review were internal memos, technical reports, white papers and proposals. Also included were journal articles published by ERIM employees which describe results of work done at ERIM. From these documents, roughly half of them were chosen as having material relevant to the SAMME requirements. In some cases, only a small portion of the document was found to be relevant. In a few cases, procedures performed at ERIM for government classified projects were known to be the same or similar to procedures described in the open literature. In these cases, the open literature was cited (e.g. Tong, 1987). Also, descriptions of some common procedures (such as level slice and histogram equalization) were included for completeness or because they were used as a basis for performance comparison with ERIM derived procedures (e.g. SIFT filtering is compared to median filtering).

Approaches and algorithms applicable to the SAMME task as found in the documents described above are identified and documented in the report. This report includes information on benefits and drawbacks of various approaches under different circumstances, when this information is available. However, this report is not a recommendation for a particular system, only a compendium of knowledge applicable to the task.

The Glitter Pageant data collection is assumed to be the source of image data for the SAMME effort. This collection includes data from forward looking infrared (FLIR), multi-spectral scanner (MSS) and electro-optical (EO), laser radar, millimeter wave (MMW), aerial photographs and synthetic aperture radar (SAR). The Have Sailor collection is especially rich in sensor types. Processing for all these types of sensor data is included in this report.

Since most raw imagery requires pre-processing for noise suppression or to remove sensor-specific distortions, the first section after this introduction (Section 2) deals with those issues. After preprocessing it is expected that the imagery would then be prepared for the automatic feature detection and extraction process which is discussed in Section 3. Once the features of interest have been identified and extracted, target features must be separated from non-target features. Methods for target discrimination are discussed in Section 4. Data from one sensor must be fused with data from the other sensors. Fusion can be done at the pixel, feature or target level, as discussed in Section 5. The candidate targets are then cued to the photointerpreter. Section 6 briefly describes target areas cueing. The effectiveness of the photointerpreter to identify targets and reject false alarms may depend on the display method and available interpretation or image enhancement aids. These issues are discussed in Section 7.

2.0 PRE-PROCESSING

Preprocessing of image data is required to condition raw images for later operations. Preprocessing includes noise suppression and contrast enhancement, as well as operations required for special classes of image data.

2.1 NOISE SUPPRESSION

Speckle is the granular look which images acquired with coherent illumination, such as synthetic aperture radar (SAR), have; it is caused by the random interference of light waves from diffusely reflecting objects when illuminated by temporally coherent light (Peterson, et al., 1988). The presence of speckle in imagery reduces the detectability of objects in the image. It also reduces the effectiveness of some computer algorithms (e.g. edge detection) designed for automatic image analysis.

2.1.1 Averaging

The most obvious way to try to suppress noise is by using a simple moving window averaging filter. Such a filter replaces the grey level value of the pixel at the center of the window with the average of the pixel values in the window. This filter is rarely used in practice because it smears boundaries or edges between contrasting regions (Miller, 1988a). Median filtering is often used instead of simple averaging. Median filtering is discussed in detail in Section 7.4.3.2.

Look-averaging is a common method of reducing speckle noise in synthetic aperture radar (SAR) imagery. It involves the noncoherent addition of multiple statistically independent images. One way to carry out look-averaging is as follows: Fourier transform the complex image and divide its square domain into four smaller squares. Each of these four squares can then be inverse-transformed to obtain four complex looks. The detected looks can be obtained by computing the magnitude of the complex looks. Finally, the average of these four detected looks can be computed (Crimmins, 1986a). The obvious drawback is that forward and reverse Fourier transforms need to be applied, so the procedure is computationally intensive.

2.1.2 Morphological Approaches

Mathematical morphology is shape recognition carried out with set theoretic operations (Sternberg, 1986). It formulates image processing algorithms into algebraic expressions whose variables are images and whose operations logically or geometrically combine images. This provides a conceptual base for processing images which is easily understood and applied. Its major benefit is that is allows algorithm developers to readily adapt algorithms to unique pattern characteristics by providing a communication channel to their spatial image character (Holmes and Sampson, 1989). Without this base it has been necessary to use signal processing techniques for images, which do not permit easy application to shape recognition (Becher, 1982).

There are six different morphological operations. They are union, intersection, complement, reflection, dilation and erosion. A dilation of one image by another is accomplished by forming the union of all translations of the origin of the first image by each of the points in the second image. Erosion is the union of all those points of the origin of the second image where it is contained completely within the first (Becher, 1982). In both cases, the second image is often referred to as a "structuring element."

2.1.2.1 Removal of Background Variation

In thresholding an image, each pixel in the image is compared to a pre-selected "threshold" value. If the pixel's grey level is above the threshold, it is replaced by one value. If it is below, it is replaced by another (section 3.1). When the grey level values of pixels of interest (target pixels) overlap those not of interest (noise pixels), then constant level thresholding will not produce the desired output. If the threshold is adjusted to detect the high level pixels, the lower level ones will be missed. If it is adjusted to obtain the low level pixels then the image will be swamped with noise.

These pitfalls are avoided by taking a morphological approach. Removal of background variation can be accomplished by eroding the image with a structuring element of a size slightly larger than the width of the target objects. The result will be an image which is black below the locus of points connecting the centers of the element. Dilation of this image produces an image black below the locus of tangents to the elements. Subtraction of this locus from the original image yields an output image which has the targets placed on a uniform background (Becher, 1982). This operation is nonlinear and does not degrade the edges.

2.1.2.2 Crimmins' Filter

The geometric (or "Crimmins") filter was designed using the mathematical morphology approach to reduce speckle in synthetic aperture radar (SAR) imagery while preserving spatial information such as edges, strong returns, etc. It is a nonlinear filter based on applying an iterative convex hulling algorithm alternately to the image and to its complement (negative of the image). It is essentially a one dimensional algorithm that is applied successively in four different directions in the two-dimensional image: horizontal, vertical and the two diagonal directions. Detailed explanations of the algorithm can be found in Crimmins (1982, 1985a,b and 1986a,b). Although the filter was designed originally for speckle reduction in SAR images, it has since found application in noise reduction for infrared (IR) imagery as well. The results of the filter are similar to the median filter (Section 7.4.3.2) in that it "whittles" down spikes while maintaining edges. But its implementation is faster than the median filter and, unlike the SIFT filter (Section 7.4.3.3) it requires no setting of thresholds. Crimmins' filter also compares well with look-averaging (Crimmins, 1986a); five iterations of Crimmins' filter produces superior results to a four-domain look averaging. Thus, Crimmins' filter can be used to produce either high quality imagery or imagery of the same quality at lower cost. Another morphological approach, the SIFT filter, is described in Section 7.4.3.3.

2.1.2.3 Convex Hulling

Gleason (1988) describes a procedure for speckle reduction in SAR imagery which uses two iterations of the complementary horizontal convex hull (CHCH) enhancement operator. The first iteration consists of one iteration of a horizontal convex hull (HCH) operator followed by an image complement transformation, followed by another CHCH operator and concluded with a second complement transformation. Subsequent CHCH iterations are executed in an identical manner on the result. The HCH operator causes the grey scale image pixels to increase in value or remain unchanged. The image complement transformation before and after the second HCH iteration in the CHCH operator effectively causes the gray scale image pixels to decrease in value or remain the same. Initial iterations of the CHCH operator cause small pixel groups with high and low values relative to their local backgrounds to be removed from the image and be replaced with values representative of the background. Two iterations of the CHCH operator remove small-scale image variations due to speckle while preserving larger-scale variations representative of the underlying scene structure, faithfully maintaining their size and shape.

2.2 CONTRAST ENHANCEMENT

Contrast enhancement serves to improve an image based on its contrast and dynamic range characteristics, typically by histogram modification. Contrast enhancement is typically applied either before or after other processing. It is discussed in detail in Section 7.4.2.

2.3 SPECIAL APPLICATIONS

2.3.1 Processing Range Data

Laser ranging imagery is acquired by scanning an infrared laser beam or a radar beam across a scene and comparing the returned signal to a local oscillator. The amplitude modulation of the received signal is phase shifted from the modulation of the transmitted signal by an amount $\Delta\phi$ which is related to the range from the sensor to the object. The value of $\Delta\phi$ is known to within 2π. In particular, $\Delta\phi = (2R/\Lambda) - m2\pi$ where R is the range, Λ is the modulation wavelength and m is an integer. The quantity $\Lambda/2$ is referred to as the ambiguity interval (Peterson, et al., 1988).

The following procedure has been shown to be effective for extracting information from laser radar imagery (ERIM Technical Report 177200-21-T, 1986): To locate the ambiguity interval the image is first filtered to remove spike noise and some texture. A first order differencing routine extracts areas where very large and distinct edges are present. More filtering and thinning occur until only the ambiguity cross-over areas are marked. At this point, vertical areas that intersect the ambiguity interval cause discontinuities in the ambiguity interval line. To make the ambiguity interval line continuous, vertical surfaces must be assigned to an ambiguity interval. A verticality image is created and the vertical objects are identified. The vertical surfaces are then assigned to the correct ambiguity interval by growing the previously determined ambiguity lines over the intersecting surfaces. The uppermost boundary of the marked vertical surface is then made part of the ambiguity interval line. At this point all of the ambiguity interval lines are marked.

An operator now positions a cursor over two calibration points. The marked points are then grown out to create isorange lines which are used in the calibration portion of the range determination. (The calibration points and isorange lines are only needed to calibrate the depression angles, which could also be supplied with the image data using appropriate hardware.)

Range determination itself consists of three steps: (1) determination of which ambiguity interval the selected point and the calibration point reside, (2) calibration of the image and determination of the number of ambiguity intervals, and (3) retrieval of the original relative range data and addition of it to the product of the number of ambiguity intervals and the ambiguity interval.

The determination of which ambiguity intervals contain the selected point and the calibration points requires that the previously extracted processed image window be searched and the ambiguity lines counted. The result of the image processing was to place contiguous isorange lines across the image at the calibration points and to make the ambiguity interval lines contiguous. Also, the processed image window is one column of the full-sized processed image which contains the target point, two calibration marks and multiple ambiguity interval marks. The search and counting routine starts counting the ambiguity intervals from the bottom of the image, corresponding to the shortest range measurements. The search and counting routines will assign an ambiguity interval number to the calibration points and the target point.

The calibration step requires that the user enter the range values of the calibration points. From this, the depression angles θ and $\Delta\theta$ can be calculated. The resultant values can then be used to determine the absolute ambiguity interval number of the target pixel.

The last step retrieves the relative range value at the target pixel location in the unprocessed image window. The absolute range value can then be calculated.

This procedure was tested on a limited number of test scenes. The performance was good but several problems were encountered. For instance, the ambiguity interval finding algorithm initially produced some false classifications. Also, edge effects occurred during the classification of vertical surfaces that intersect the ambiguity interval cross-over. The edge effects produced slightly inaccurate placement of the ambiguity interval line. A more sophisticated classification algorithm would be required if this were to pose a major problem. The algorithm also requires an area of open ground in front of the target so that good ambiguity interval lines can be determined. Further work would have to be done with images acquired in more cluttered terrain.

3.0 FEATURE DETECTION AND EXTRACTION

The purpose of feature detection is to locate potential or candidate targets. Raw or pre-processed data is evaluated to find features of interest. In feature extraction, the image is partitioned so that regions of pixels associated with features (potential targets) are separated from regions without features of interest (background). Techniques vary from simple level slicing to complex morphological and statistical approaches.

3.1 LEVEL SLICE

The simplest approach is to threshold the imagery. While not usually used alone, this thresholding or level slicing is often part of a more complex algorithm (e.g. SLOR, Section 3.3.1).

Binary operations can be used to process an input image into a high contrast output image consisting of only two grey levels, usually black and white. Each pixel grey level in the image is compared to a preselected "threshold" value. If the pixel's grey level is above the threshold, it is replaced by white. If it is below, it is replaced by black. This process is useful whenever the pixels of interest are known to be above a certain grey level, since then all the other pixels, not of interest, can be suppressed. The drawback is, of course, that often the pixels of interest and the background pixels have common or overlapping grey levels. Setting the threshold too high will eliminate some of the pixels of interest, setting it too low will result in background (or noise) pixels being included, often to the extent of overwhelming the target pixels.

One way around the drawback of the binary level slice is a color level slice, where several thresholds are set and pixels within ranges of values are assigned to a particular color (or grey level if a color monitor is not available). Many or few thresholds may be set, depending on the application. The drawback here is that unless the imagery is very consistent and there is sufficient a priori knowledge to meaningfully pre-set the thresholds, this method can require extensive anu time consuming interaction by a photointerpreter. Even then, it may be that some target pixels will be eliminated and background pixels enhanced for the same reason as in the binary operation, and that the only way to retain all the information would be to set the ranges so small as to nearly be equivalent to the original image. Finally, color slice imagery requires quite a bit more training to interpret and understand.

3.2 SPECTRAL METHODS

Spectral methods have been used extensively in remote sensing. The essence of remote sensing involves discriminating between various materials of interest and their backgrounds based on the radiation received by the sensor. Because materials have unique spectral characteristics (signatures), multispectral remote sensing techniques can make use of observations in more than one wavelength interval to discriminate between classes of materials. In conventional multispectral recognition, the total area of each ground material is measured by identifying the material in each ground area (pixel) covered by one resolution element of the sensor. The total area covered is found by adding up the pixels identified with that material. If almost every pixel in the ground scene contains just one of the possible materials, this technique provides adequate estimates of coverage.

Roller et al. (1981) developed and implemented a procedure for classifying crop types which may have applications to SAMME when multispectral data is available. Several data sets from different times are initially screened. Those data sets with heavy cloud cover, heavy haze or other adverse effects are deleted. The selected data is normalized to account the effects of light haze, varying sun angle and sensor calibration. Features are extracted from the normalized data using the Tassled-Cap transform (Crist and Cicone, 1984), a generalization of the Principal Components transform. Temporal pattern classes are then extracted with are based on the pattern of vegetation development during the course of a growing season. The temporal pattern classes are then used to automatically assign all the pixels to major crop types. Blobs are identified by defining field-like targets of single crops. Once the blobs are defined they are separated into two groups according to size: "big blobs" which consist of all blobs that have at least one pixel in their interiors and "little blobs" which have no interiors. Only the big blobs are used as candidate labeling targets. This separation is carried out in order to isolate mixed pixels and very small fields which prove to be poor labeling targets. Each blob, big or small, is then assigned to one of the crop group stratum according to the vegetative development pattern it exhibits. A sample of the big blobs are then used in the unsupervised clustering algorithm. Each sampled blob is checked for mixture, i.e. the probability that it contains more than one crop type. Blobs not flagged as mixed are passed into the automatic labeler. Mixed blobs are fed through the labeler one pixel at a time. If most of the pixels in the mixed blob can be assigned a label, the proportion of the total number of pixels in each label category is determined and a label reflecting the observed mix is assigned to the blob. If the number of pixels that can be labeled is insufficient to justify assigning a label, then the blob is passed to a human analyst who is provided with a number of labeling aids.

Spectral discrimination would be much simpler if the spectral content of the received radiation remained constant, but it does not. Throughout the day the sun changes position and the properties of the atmosphere change. Thomson and Sadowski (1975), for example, showed that the major effect of changes in path radiance on recognition accuracy was with classifying high reflectance materials. They found that, although variations in path radiance also caused changes in the signatures of low reflectance objects, the decision regions for these materials were large enough to accommodate the changes with little if any change in accuracy.

Studies have been made (Malila, et al., 1971 and 1972) for improving multispectral discrimination techniques in the face of changing conditions and of extending their capability. A unified radiative transfer model was developed and parametric calculations of irradiance, path radiance, and transmittance were made. Model calculations were then used in a statistical simulation of a complete recognition system on a problem that demonstrated the deleterious effects of haze (path radiance) on recognition results. Finally, the conventional assumption of multivariate normal signal distributions was examined and found to be untrue according to a χ^2 test applied to selected empirical data. Other linear decision rules have been shown to be superior to the popular quadratic decision rule based on the Gaussian assumption (Crane et al., 1973). The so-called "best linear decision rule" was shown to have a smaller probability of misclassification for equal processing times, or a shorter processing time for equal probabilities of misclassification on test fields.

Nalepka and Morgenstern (1973) also attempted to increase classification accuracy by correcting for path radiance effects. Attempts were made to estimate the path radiance effects using an ERIM radiative transfer model. Because of problems associated with the calibration of the data and with model parameter specifications, this approach was unsuccessful. A second approach was devised in which the smallest signals at each scan angle were used as an estimate of path radiance. Results of classifying data modified in this manner were inconclusive.

Many multispectral classification rules are based on information from one pixel at a time. Algorithms have been developed which increase the accuracy of multispectral recognition with only a small increase in cost by using classification rules that use data from groups of pixels, such as the "nine-point classification" algorithms (Richardson and Gleason, 1975) developed for NASA. This set of rules determines what ground cover category to assign to a pixel based on data from that pixel and from its eight immediate neighbors. Such rules are applicable only when a pixel is likely to represent the same material as its neighbors. This approach can be used to categorize surface types which can then be used in target areas cueing.

In sampled imagery, many of the pixels will have received radiation from several objects and so will not be representative of a pure spectral signature. Conventional algorithms will not be adequate for classifying such pixels. Modifications to classification algorithms have been made (Horwitz, et al., 1974 and Horwitz, et al., 1975) which improve the basic proportion estimation algorithm as well as improving alien object detection procedures. The modifications are based on determining which pixels are likely to be on a boundary and estimating the proportion of classes within the pixel. A simplified signature set analysis scheme was introduced for determining the adequacy of signature set geometry for satisfactory proportion estimation. In the study, averaging procedures used in conjunction with the mixtures algorithm were examined theoretically and applied to artificially generated multispectral data. However, experiments conducted to find a suitable procedure for setting the alien object threshold yielded little definitive results.

When multitemporal image data is registered and classified, errors in classification can result from misregistration of the data. A capability was developed (Malila, et al., 1975) to compute probabilities of classification for any signature distribution with respect to the optimum linear decision surface defined by any given pair of signatures. This capability was used to compute probabilities of detection and false alarm for a variety of distributions for misregistered pixels. From these studies recommendations for improvements to the approach were suggested.

Thermal and reflective signals are produced by different physical mechanisms. A thermal signal results from self-emission that depends on the temperature and emittance of the material, whereas a reflective signal depends on the incident radiation and reflective characteristics of the material. This difference in the origin of the signals can be used to improve discrimination. However, this difference can also lead to misclassification. For instance, the temperature of a material on the ground depends on physiological, physical and meteorological factors often having little, if any, influence on the reflectance of the material. Also, the thermal signatures depend not only on the existing environmental conditions but also to some extent on the past history of such conditions. The length of history that is important depends on the type of surface. Finally, the way in which spectral signatures change with time and distance from the training data will differ in the two spectral regions because of the different physical processes involved. Malila, Crane and Richardson (1973) have investigated some of these issues and developed a recognition processing procedure. Sample maps of urban and rural areas were made using the procedure, but additional analysis was recommended to account for atmospheric effects and bidirectional reflectance properties of surface materials.

3.3 SPATIAL METHODS

If the object of interest can be described, then a spatial automatic target recognition system can be made to work (Brown and Swonger, 1988). For the most part, the data processing required is driven by the variability in responses to a target that sensors may have. Of course, if responses to targets and responses to clutter objects overlap considerably, performance is limited.

3.3.1 Morphological Approach

The following algorithm is an example of an approach for feature detection which uses mathematical morphology. The target detection and extraction problem can be approached as a series of levels, as shown below (Rauchmiller, 1989):

Level 1 <u>Classify Frame</u>

1 = Frame contains targets

0 = Frame is void of targets

Level 2   Classify Pixels in Frame

1 = Pixel is of interest

0 = Pixel is of no interest

Note that if such an approach is used it is imperative that level 1 have nearly 100% detection probability. As the algorithm progresses through the levels, there will be ample opportunity to eliminate false alarms. But any targets that are missed at level 1 will never be recoverable. Thus, level 1 processing must tolerate a high false alarm rate.

The Slice and Open Residue (SLOR) algorithm processes SAR imagery through level 2 (Rauchmiller, 1989). It locates small high contrast targets. It does this by processing each frame in two different ways, then applying a logical AND to the output of each processed image to produce an output image which is a set of points where the candidate targets are located. One of the processing paths involves first a level slice of the imagery which segments the data into a binary image (Section 3.1). The other processing path first iterates on the Crimmins' filter (Section 2.1.2.2) which suppresses speckle and low frequency noise. This filtered image is then subtracted from the original, producing a "residue" image. Image opening and closing are then performed using a conical structuring element, and the result segmented into a binary image. The two binary images from the two different processing paths are then logically ANDed, so that only blobs which overlap are kept. These blobs then undergo a point reduction to create a final output image which is a set of points where the candidate targets are located. The major drawback of this method is the setting of thresholds. If set too high, many target pixels will be eliminated. If set too low, too many noise (non-target) pixels will be introduced. Thus, target pixels which are of low intensity relative to the highest intensity target pixels will be impossible to detect.

An approach for extracting roads is presented in Gleason (1988). It should also be applicable to extracting any target of known size and shape. First the image is enhanced using two iterations of the complementary horizontal convex hull operator (Section 2.1.2.3). Then regions are removed if they do not satisfy specified size criteria. This is done by eroding (Section 2.1.2) the enhanced and complemented image four times using all eight nearest neighbors on each iteration. The four erosion iterations cause each pixel in the image to be replaced by the minimum of a 9-by-9 array centered about it. Small regions are completely removed. The shapes of the large regions are recovered by an iterative grey scale spanning (GSPAN) operator. On each iteration the image is dilated (Section 2.1.2) over a 3-by-3 neighborhood (each pixel is replaced by the maximum of itself and its eight nearest neighbors) and the result then intersected with a mask image (each pixel in the dilated image is replaced by the minimum of itself and the corresponding mask image pixel). The resultant image is then used in the next GSPAN operation. Then the image can be level sliced (Section 3.1) at thresholds chosen to cover a dynamic range of typical target pixels and the resultant binary images are combined by a differencing operator. This leaves only the target sized regions which can then be extracted as, say, eight-way connected components.

3.3.2 A.I. Approaches

Features may also be extracted and manipulated at a symbolic level. The image region extraction system AGGREGATE (Walters, 1987) does just that. The user specifies an input look-up table identifying pixels of interest in the image. The image is runlength encoded from left to right, top to bottom. Runlength attributes include location and image intensity information. After runlength extraction, the direct (image) spatial relationships between features is lost. But runlengths do retain their coordinates, and the ordering of the runlength vector is such to make runlength nearness measurement efficient.

Runlength merging is then performed for chaining together runlengths that are near each other in the image and have the same input look-up table value. There is a one-to-one correspondence between run-length chains and regions. A region is constructed for each runlength chain. Proximity sets, sets of regions that are near to one another, can then be extracted. The implementation of the construction of region sets is analogous to that of runlength merging.

Spectral or spatial filters can be applied to the regions followed by an extractor, with allowances for pixel gaps in components. The extractor writes out region files containing an oriented bounding rectangular box for each region. The box has position, length, width and orientation angle attributes. The box position is in image coordinates. Thus, if this region list is to be compared to a list from some other image, it is essential that the images first be fixed to a common coordinate system.

This is perhaps the major drawback of this method for SAMME applications: Imagery from different sensors may very well have different ground projected resolutions or for other reasons be very difficult to project to a common image coordinate system. (The obvious modification would be to fix all the images to a geographic coordinate system.) A less critical drawback is the use of rectangles to identify the regions of interest. The choice to do this in AGGREGATE was made because the targets of interest for this work were roughly rectangular in shape. If the targets of interest were, say, U-shaped, then using rectangular boxes to represent them would not be effective.

3.3.3 Dynamic Programming Approach

Gleason (1988) describes a dynamic programming approach for road extraction from SAR imagery which may have application to the detection and extraction of other types of objects. The first step in the process is to create horizontal and vertical contrast images using 41-by-1 and 1-by-41 median filters. The dynamic programming step then begins by creating "cost" images associated with the four scan directions. These cost images will be used in a traceback step to find dark paths (candidate roads) through the image. The horizontal and vertical contrast images are scanned, the scanning direction tells how states and stages relate to the rows and columns of the image. Associated with each state and stage are the values of three functions: the input value, the total cost of the best path from that state and stage, and the length of the best path from that state and stage. The cost and length functions are generated during the dynamic programming process. The value of the cost function at a particular state and stage is the total cost of the best path leading back from it, where "best" in defined in terms of a minimum. The dynamic programming process starts with the input values and fills the cost and length functions starting with the first stage and ending with the last. At the end of the dynamic programming, an average cost array is formed by dividing each total cost value by its associated path length.

The next step in the road extraction process is the traceback step. This process generates actual paths through the image which, when combined, make up the edges in the graph structure. The traceback step starts at the last stage of the (average) cost array looking for places to start a traceback path. If an average cost value is the smallest value within a window in the last stage, then a traceback path is started from that point. Traceback is effected by progressing from the last stage to each previous stage until the first stage is reached, choosing the next point with the minimum average cost value. Once the paths from the four scans are generated, they are combined and translated into a symbolic graph data structure. The paths are combined into one image, and end points and crossing points are found. These become the vertices of the graph structure. The paths between the vertices form the edges of the graph and are stored as lists of points. The resulting graph structure is stored symbolically.

The next step is to measure attributes of the edges of the graph. The measured attributes are length, straightness and contrast. Lists of seed edges are then produced based on these attributes. These seeds are extended by examining edges leading from vertices which are endpoints of seed edges. The final set of edges constitutes the output of the road extraction process.

This dynamic programming segmenter was shown to be far superior to conventional edge detection methods in the presence of speckle. In particular, it had a lower false alarm rate and was better at following roads through areas where they were indistinct.

However, several problems were found with the approach. Bends in roads which were more than 45 degrees from the vertical were missed in the vertical scan and, if they were not picked up well in the horizontal scan, would not be detected at all. Several possible solutions to this problem were proposed: the range over which the minimum cost function was taken could be increased, a second pass of horizontal dynamic programming could be used, or another kind of dynamic programming which moves in radiating waves from a given starting point could be used. The approach as it currently stands has no method for rejecting seed segments. This resulted in extraction of non-road features, such as tree lines. Higher level reasoning based on the extendability and connectedness of the candidate road segments was proposed as a possible method for dealing with this problem.

3.3.4 Model-Based Approach

Zelnio (1986) suggests the following detection and segmentation algorithm for use on SAR imagery. The size and shape of the object of interest is assumed to be known. A double window is chosen where the "object window" is the size and shape of the illuminated edge of the object and the "clutter window" is the direction of the radar. (This is motivated by the fact that the illuminated edge is likely to contain areas of significant scattering while the adjacent area is likely to contain non-shadowed clutter.) The double window template is translated to each pixel position in the image and a statistical measure of texture is computed over the area of the object window and compared to the same measure computed over the clutter window area. Texture is used because it differentiates well between the coherent scattering resulting from man-made objects and the diffuse scattering from natural objects. The window is rotated and the measure recomputed at each aspect angle increment. At each translation the value of the maximum texture difference corresponding to the best aligned aspect angle is saved. The detection decision process first thresholds on the pixels having acceptably high values then finds local maxima corresponding to the precise object location.

The main benefit of this detection process is that it also doubles as a segmenter since an aligned template at the proper location is the detection result. If objects of different size or shape are expected the procedure can be repeated with different templates. The one with the largest difference in texture measure is the one chosen for segmentation. Of course the obvious drawback of this method is that it is incredibly computation intensive.

3.3.5 Background Cancellation

Background cancellation is another common approach (Kryskowski, 1989; Kuschel, 1988). A "background" image is produced using several iterations of a speckle reduction algorithm (Section 2.1) followed by smoothing. The original image is compared to this background image. Objects are detected by taking the ratio of the scaled version of the original with the background image. This yields an image with bright detections on a dark background. One of the benefits of this method is that it is locally adaptive since brighter background areas require brighter differences to be significant.

3.3.6 Region Growing

Region growing is a segmentation method based on enlarging regions about a "seed", a pixel representative of the region of interest. Typically this is done based on grey level thresholds. For SAR imagery it can be done based on spatial relationships and illumination direction (Kuschel, 1988). The lowest grey levels are assumed to be shadows and the brightest are chosen as the seeds for the bright returns. Seed grey levels based on radar cross section calibration information are also chosen for natural features such as grass and trees. The region growing considers each grey level in turn, comparing each pixel to its neighbors and their region membership. It assigns each pixel individually to one of its neighbor regions if certain spatial relationships hold. Since each pixel is tested individually, different regions (e.g. grass and trees) may contain pixels of the same grey level. This is exactly opposite to thresholding (Section 3.1). Spatial reasoning can then be used to reinforce the initial segmentation.

This method can also be used for change detection, although to do so requires stable image geometry, uniform illumination and accurate radar cross section calibration. The method is fairly computation intensive, but can be applied to spatial compression imagery with a large reduction in processing time at a minimal cost in the loss of small regions.

4.0 TARGET DISCRIMINATION

Target discrimination takes the regions resulting from the feature detection and extraction phase and computes a set of salient characteristics for candidate target regions which can be used to discriminate between valid targets and non-targets. If more than one type of target is present in the image, this step should also partition the target list into appropriate classes. One way to conceptualize the discrimination phase is that it appends onto the two levels of the detection and extraction phase three more levels (Rauchmiller, 1989):

Level 3   Classify Interesting Pixels
- 1 = Pixel is of primary target
- 2 = Pixel is of secondary target
- 3 = Pixel is of tertiary target
- . . . .
- n = Pixel is of n^{th} order target

Level 4   Layers of Orientation Processing

Level 5   Aim-point Selection
(i.e. locate specific part of target)

4.1 CORRELATION TECHNIQUES

Correlation techniques calculate some measure of similarity for the intensity values of portions of one image against relative shifts of the other (often called a "template"), the best match being based on some match quality criterion. Although this is generically referred to as "correlation", it may not be correlation in the strict mathematical sense. Correlation type techniques provide inherent noise suppression and can result in good matches even with significant amounts of noise, as long as the relative geometric distortion between image and template is not great. The benefits and trade-offs of the various correlation-type approaches are discussed in Wood (1989).

4.2 STATISTICAL PROCEDURES

Statistical procedures may also be used for target (or image "chip") matching. Statistical procedures are those which choose among alternatives based on probabilistic models of the alternatives. For example, one could use quotient statistics (Miller, 1988b). Briefly, if X_i and Y_i are the returns from corresponding pixels in two target images, then the images are declared the same if the spread of the ratio X_i to Y_i for $i = 1,2,\ldots N$ lies in some pre-selected interval. This spread could be defined by the maximum and minimum ratios, or by some other combination of order statistics. The benefit of this approach is that computations with these types of quotients are very simple to perform. If there is a constant bias between the two image chips, (for example, if different gains were set for each image), then the approach can be easily modified to take this into account. The procedure can also be easily modified to take into account any speckle reduction that has already taken place simply by changing the appropriate parameters. The procedure is described, along with the mentioned modifications, in detail in Miller (1988b).

Miller (1988a) also proposes a Likelihood Ratio Detector which computes a test statistic which is defined as the average of the received power from a set of contiguous resolution cells (the target template) being tested for the presence of a target, divided by the average of the received power from another set of resolution cells (surrounding the target template) that are used to characterize the local clutter background in the vicinity of the target template. This procedure can be implemented in a way that maintains a constant false alarm rate, which is desirable for maintaining a constant and predictable processing load. The procedure requires a user-specified threshold. Pre-filtering of the data (using moving window averaging) helps to improve detection performance. Performance on imagery prefiltered with a median or other nonlinear filter was not discussed.

Statistical procedures may also be used for moving target detection in SAR imagery. Target motion in the cross-range direction causes streaks or smears on a SAR image. Statistical detection procedures can be used to detect these streaks. For example, Miller (1988c) describes such a procedure which is based on thresholding the quotient of the average power in a template matched to the expected streak size and the average background clutter power in the vicinity of the template. Miller illustrates the method in a worst case clutter analysis that shows the method to have promise; but the consequences of various levels of streak false alarm remain to be examined in terms of image processing requirements and follow-on discriminants that might be required.

4.3 SYMBOLIC ALGORITHMS

Symbolic algorithms require symbolic representations of features such as edges, lines, vertices of line intersections, shapes and so on. A match quality criterion (such as distance) is used to compare map and image symbolic data sets. The symbolic data has a much lower dimensionality than pixel-based data resulting in reduced computation for these methods compared to the correlation-type methods. On the other hand, symbolic matching strongly relies on noise-free feature extraction. Thus symbolic techniques will fail on noisy images unless good preprocessing for noise suppression is applied.

Gleason (1988) describes a method for the symbolic identification of roads from SAR imagery, which should be extendable to identifying other objects in images. Symbolic domain operators generate a sequence of progressively higher level hypotheses about roads using both bottom-up and top-down control strategies. Global road hypotheses are asserted for colinear groups of candidate road segments already identified and extracted (Section 3.3.3) from the imagery. This process is directed by a bottom-up control strategy where computational cost is minimized through the use of heuristics which cause attention to be focused on those groups of candidate road segments that hold the most promise for supporting a global hypothesis. Heuristics also play a vital role in avoiding the formation of multiple redundant global road hypotheses. Global road hypothesis formation is an iterative process driven by a set if global road seeds, sets containing candidate road segments with lengths exceeding a minimum seed length parameter.

Extended road segment hypotheses are derived from the global hypotheses based on the spatial properties and values of the original image pixels which fall under the global hypotheses. The extended road segment hypothesis formation process has two primary purposes: to discriminate between roads and backgrounds based on a top-down focused analysis and to provide a more accurate delineation of the spatial extent of roads represented by edge-to-edge global hypotheses. In this process, gaps and segments are deleted iteratively until all have length exceeding a maximum detection parameter.

This method was shown to be a viable approach for road identification and extraction from SAR imagery. The candidate road segment hypothesis was found to be the most critical factor controlling the success of the procedure. Some roads yielded a dense set of candidate segments and posed no difficulty to the subsequent processing. Others, however, yielded only a minimum set of segments. Correct global road hypotheses could be asserted based on such minimal evidence but only at the expense of a higher false alarm rate. Improved low level feature extraction methods would be helpful in reducing the number of regions extracted as input to the symbolic operators. Also, it was felt that region extraction and symbolic processing could be integrated more tightly so that the higher level hypothesis formation results could be used in directing lower level processes. Finally, the higher level processes could have been developed to exploit expected road pattern interrelationships.

4.4 EXAMPLES

4.4.1 Finding Tanks in Downlooking 3D Range Data

A target cueing and classification system was developed at ERIM for automatic tank recognition in 3D range data (Holmes, 1989). Needle shaped structuring elements (Section 2.1.2) oriented along the scan lines were first applied to normalize the background. When rolling terrain was present it was necessary to divide the image into separate regions and apply differently oriented needle shaped elements to the different regions. Then relative range level slices (Section 3.1) were used to obtain "body" profiles and "turret" profiles. First the body profiles then the turret profiles were tested by alternatively dilating and eroding (Section 2.1.2) with the appropriate structuring elements. Then only locations which weren't too close to tall objects were accepted. Finally, cue circles were drawn around the tanks on the input image. This study was done on simulated data so its effectiveness on real data is not known.

4.4.2 Mine Field Detection

Morita et al. (1979) report a method for the detection of mine fields which does not require that the detection algorithm actually identify individual mines by their characteristic signatures, but simply extracts features which might be mines. Large numbers of false alarms result. The mines were expected to be deployed in a particular pattern, so the algorithm looks for the pattern among the detections. If some of the detections do fit the expected array pattern approximately, the probable locations of undetected mines can be found from any missing parts to the expected pattern. Morita et al. did not test their algorithm against real backgrounds, so it is unknown how it would perform in the presence of other types of arrays, such as fence posts, hay stacks or orchard tree stumps.

5.0 DATA FUSION

Data fusion is the ability to combine data from different sources in a meaningful way. Often different types of imagery can be combined in such a way that the strengths of each are exploited. For instance, one sensor may have the range of spectral bands necessary to spectrally but not spatially distinguish features, while another may have fine spatial resolution but in only one spectral band. Similarly, a nadir pointing satellite image might yield both spectral and spatial information but not elevation information, so fusing the image with cartographic information would help to resolve issues which were terrain related, such as trafficability. Knowledge of surface material composition such as soil type and moisture content would also aid in trafficability analysis. Finally, very fine resolution imagery might show potential targets very well, but without enough context to make accurate targeting decisions. In such cases, the ability to place the target in context would be important and might require fusing imagery from two sources.

Fusion from multi-source data is also beneficial in that it provides for robust operational performance, increased confidence, reduced ambiguity, improved reliability in terms of lowering false alarm rates and raising detection rates, and improved classification. If the sensors are colocated or nearly so, then the processing and fusion leading to target recognition and identification will be relatively easy. Alternatively, the raw data from the sensors could be processed at the sensor site up to the feature or target detection level and transmitted to a common processor for correlation and fusion with other reports.

5.1 FUSION AT THE PIXEL LEVEL

In order to fuse imagery from different sensors or times at the pixel level, or to fuse cartographic data with images, it is necessary to first geometrically align the images or image and map so that features in one correspond in location to features in the other. This will most likely involve rather complex mathematical models, careful attention to detail and a great deal of computer processing. Issues likely to be encountered in pixel level image fusion are discussed in this section.

5.1.1 Geometric Correction

Some geometric "warping" or transformation must be applied to remotely sensed image data to remove systematic distortions which are introduced by the sensor and its platform. Most likely, the images will also be required to be registered to a specific map projection. This is called "georeferencing" and its goal is to change the geometry of the image so that features are found in the same location in image and map: that is, so that each pixel in the image can be associated with a unique latitude and longitude. Images acquired at different times or by different sensors are registered to a common map projection in order to facilitate meaningful comparisons. Georeferenced images from different sources can be readily used for change detection or for comparisons with other sources, including non-image sources such as maps. Alternatively, images could (and often are) referenced to one another. However, referencing to a pre-selected map projection is a more general approach. (Alternatively, maps could be distorted to fit the image geometry, but different images corrected in this way could not be easily compared.)

Numerous sources contribute to the distortion of unprocessed image data which make georeferencing necessary. These distortions are both sensor induced and platform induced. The sensor induced distortions are consistent over time and can be modeled in a systematic manner. Among these may be included panoramic distortion, lens distortion and any nonlinear motion of an oscillating mirror. Such distortions are well described in engineering design and test documentation (Nordman and Wood, 1988).

The platform induced distortions must be modeled based on information which changes continuously. Among these distortions are attitude (yaw, roll and pitch), altitude, heading and velocity. These platform induced distortions, while well understood, do not occur to the same degree in each image and may even change during the acquisition of a single image. The latter is especially true with airborne-acquired data. Thus, they need to be modeled based on auxiliary information that is usually supplied with each image data set. The accuracy with which they can be determined depends on the accuracy of the auxiliary information available. In theory, these distortions could be corrected with arbitrary accuracy.

5.1.1.1 Registration of Data

Both these sensor and platform induced distortions can be combined to produce an overall geometric model representing the transformation required to estimate pixel locations in an image with desirable geometric characteristics. Since it is unlikely that the auxiliary information describing the platform characteristics is sufficient for arbitrary accuracy, an iterative approach is sometimes taken which allows refinements of the model. Alternatively, ground control points (points whose locations are known precisely in both image and map) can be used to refine the model.

An analyst can locate the ground control points in both image and map. Typically ground control points are such things as road or railway intersections. Natural features such as a bend in a river or the tip of a peninsula could also be used, but since such features are likely change with season it is not advisable to use them as control points unless there are no other options. The analyst assigns to each control point a latitude, longitude and elevation. Accurate elevation information is especially necessary in regions of high terrain relief, especially with images taken from other than nadir views, as failing to take elevation into account may result in significant displacement errors.

Alternatively, if enough information is known about an area ahead of time, it is possible to have a library of control point "chips", small subsections of imagery containing features of interest which can be used in a automatic processor to find the corresponding features in the new image (Wood and Sellman, 1987). Candidate chips are selected from the library and can be roughly assigned to neighborhoods in the image by estimating their locations from the systematic correction coefficients. Some subset of the candidate chips is usually chosen to then be automatically correlated with the image. The results of the correlation serve to refine the geometric transformation which is based solely on the system model. The automatic selection and location of control points is, of course, much faster than manual selection. However, the exact method of locating the points must be carefully selected and tested. Simple correlation can actually be very ineffective. A complete description of the pros and cons of various correlation-type methods for image registration can be found in Wood (1989).

An alternate approach to correlation-type techniques would be to maintain two lists of edges (or even features) at various orientations and their locations, one list from the current data acquisition and one from a previous acquisition. The locations of these edges in the second acquisition could be estimated using auxiliary information, such as Inertial Navigation System (INS) data. Then the edges could be matched by finding those entries in corresponding lists which are within a certain distance based on known changes in location of the sensor relative to the targets (Wessling, 1989).

Whether the control point pairs are selected manually or automatically, a well distributed set should be obtained and blunder checks made. Then, for a given set of control point pairs, a least squares algorithm can be used to compute the refinement to the model based on minimizing error. This error can then be reduced by rejecting or adding control points. Control points can be iteratively added and subtracted until the error reaches an acceptable limit. The image is then resampled to the desired map projection.

5.1.1.2 Choice of Map Projection

Map projections are necessary to depict on a flat surface large areas of the globe. This cannot be achieved without distortion of scale, shape or both. Hundreds of map projections exist and the choice of one particular projection must be based on compromise suited to the application. With only one exception, all of the families of projections make use of non-linear mapping between geographic coordinates (latitude and longitude) and a flat rectangular grid in a way that attempts to minimize one sort of distortion at the expense of another (Dye, 1988a). For some of these (such as Universal Transverse Mercator) the mapping parameters must be changed with geographic location in order to map the entire globe, while with others (such as the Mercator) it is impossible to map the entire glob with a finite amount of grid space. The only exception to this is the Cylindrical Equidistant projection, in which the geographic coordinates are simply taken to be rectangular coordinates. The drawback is that the distortion is severe near the poles.

If the areas of interest are geographically small, however, it may be preferable to use a map projection which preserves area or shape. Equal area (or homolographic) projections include Alber's Equal-Area Conic projection, the Cylindrical Equal-Area projection, the Lambert Azimthal Equal-Area projection and the Sinusoidal projection. Shape preserving (or conformal) projections include the Bipolar Oblique Conic Conformal projection, the Lambert Conformal Conic projection, the Mercator, Oblique Mercator and Oblique Space Mercator projections, the Stereographic projection and the Transverse Mercator projection.

5.1.1.3 Resampling

Once the model is finalized it is used to find pixel locations in the original image from which to determine the gray level values to assign to the pixels in the corrected image. Typically, these locations will be intermediate to the pixels in the original grid, that is, they will not be at integer locations. The locations will always be somewhere between four valid pixel locations, so that some interpolation method can be used. However, interpolation acts as an added convolution, or blurring, of the image, resulting in a loss of resolution (Schowengerdt, Park and Gray, 1984). Each time an image is resampled it is blurred further. Thus, while it is possible to resample an image many times, once for sensor distortion, once for platform distortion and again for control point refinement, for the best product all the distortions should be accounted for in one georeferencing formula and the imagery should be resampled only once.

Georeferencing thus takes place in two steps: determination of a projection function which will transform the geometry of the image to that of the map, thus reconstructing a (model) continuous image, and interpolation of the image gray levels to determine the correct gray level values to assign to the pixels.

A bivariate polynomial surface may be used to model the geometry of the scene. Often the image is broken up into quadrilateral or triangular patches and a separate polynomial surface fit to each subimage. Alternatively, a single global polynomial may be used if the distortions are sufficiently smooth. This projection function defines a continuous image estimate or "model." Once that model has been determined, it can be evaluated at the desired grid locations. It is necessary to determine the "best" gray levels to assign to the new pixel locations, given the set of samples from the original image. The problem then is to determine an appropriate resampling function.

Most resampling methods are attempts to reconstruct the continuous image before it was sampled (or "discretized"), and so to retrieve the grey levels at intermediate locations. Such methods are characterized by interpolation functions which "weigh" adjacent image pixels and reconstruct the image between the original samples (Schowengerdt, et al., 1984). These methods should not be confused with restoration, the goal of which is to estimate the original scene radiance rather than the continuous image (Simon, 1975; Park and Schowengerdt, 1983).

5.1.1.3.1 Nearest Neighbor. A straightforward approach to image reconstruction would be simply to choose as the new grey level value that of the pixel nearest the one being interpolated. This "nearest neighbor" approach can be implemented quickly and efficiently. However, it can lead to position errors of up to ±1/2 pixel (Simon, 1975) and results in a blocky appearance in the final image. Additional problems are encountered when attempting to register two images resampled using nearest neighbors. Registration of details in some regions can be perfect, while severe misregistration can occur in other regions (Billingsley, 1983). This would result in significantly degraded change detection derived from the two images (Simon, 1975).

On the other hand, nearest neighbor is the only interpolation method which preserves the radiometric fidelity of the image and as a result introduces no new spectral classes (Billingsley, 1983). For some applications radiometry is a major consideration, such as feature detection and identification in a single image. For these applications nearest neighbor would be the interpolation method of choice.

5.1.1.3.2 Bilinear Interpolation. A smoother resampled image can be generated when adjacent pixels are allowed to influence the estimation of intermediate pixel values. One such approach is to assign a grey level value which is a bilinear interpolation of the nearest four grey levels. This method uses the distances to the four nearest pixels as weighting factors to interpolate the new grey level. The blockiness apparent in the nearest neighbor resampled image does not appear in a bilinearly interpolated image, and mean squared resampling error is improved by about 1/4 over nearest neighbor (Shlien, 1979). However, the method requires significantly more computer time and suffers some loss of detail due to attenuation of the higher spatial frequencies, which results in a slightly blurred appearance (Shlien, 1979).

5.1.1.3.3 Sinc Function. For an image to be faithfully reconstructed, all of its spatial frequency components must be reproduced without distortion and no new frequencies introduced. Thus, an ideal interpolation method would have a rectangular frequency response, flat up to the Nyquist frequency and zero thereafter (Shlien, 1979). In the spatial domain, this would correspond to a sinc (sin[x]/x) function. Use of a sinc function as an interpolator thus assumes that the function describing the continuous image intensity distribution is band limited (the image energy spectrum is zero for all frequencies greater than some cutoff frequency) and sufficiently sampled (at a "Nyquist" sampling rate of at least twice per period for the highest frequency present in the image). Only then can the image intensity distribution be reconstructed exactly from convolution with a sinc function.

It is impossible to implement this ideal interpolator numerically because of its infinite extent. Moreover, the image intensity distribution is known only over a finite extent. The function can be truncated, but since the sinc function decays slowly to zero it would still require a very long span to minimize ringing in the reconstructed image which would arise from the inevitable slope discontinuities at the truncation points (Shlien, 1979; Park and Schowengerdt, 1983).

5.1.1.3.4 Cubic Convolution. The most popular reconstruction method is a piecewise cubic polynomial of limited extent which approximates the sinc function called "cubic convolution." It is smooth and spatially limited and has no slope discontinuities at the endpoints.

As with bilinearly interpolated imagery, the blockiness apparent in the nearest neighbor resampled image does not appear in the cubically convolved image. But again the cubically convolved image has a somewhat blurred appearance. Also of particular concern is overshoot at boundaries between contrasting regions in the image.

Both of these latter drawbacks can be reduced by applying cubic convolution in parametric form (Park and Schowengerdt, 1983) and using the parameter to minimize overshoot and radiometric error. This parameter, α, is the slope of the interpolation function one sample length from the center. Standard cubic convolution is implemented with $\alpha = -1$.

However, Park and Schowengerdt (1983) have shown that this is not the optimum value for most images. They showed that if the energy spectrum of an image is known or can be estimated, then a value of α can be chosen to minimize the radiometric error for that image, and that α = -0.5 represents a good default value for general purpose implementation. As a matter of fact, for images dominated by low frequencies, bilinear interpolation is actually superior to cubic convolution unless α = -0.5. Parametric cubic convolution with α = -0.5 yields a smaller value for the mean squared radiometric error than either bilinear interpolation or nearest neighbor. However, bilinear interpolation yields less error for the standard α = -1 for any image which has all its energy at frequencies below 0.12 cyles per sample interval. For images dominated by edges and with a sampling rate of one, Park and Schowengerdt showed that the optimal value of α is approximately -2/3.

The blurring and edge overshoot can only be reduced by adjusting the parameter α, they cannot be eliminated. (When α = 0, cubic convolution reduces to a smooth approximation to bilinear interpolation and the overshoot is eliminated but the blurring is not and a stair-stepping artifact is introduced (Park and Schowengerdt, 1983).) In any case, whatever the value chosen for α, cubic convolution is still an attempt to reconstruct the original image before sampling, not to reconstruct the scene from which the image was acquired.

5.1.1.3.5 Restoration. All of the interpolators discussed so far act as an added convolution, or blurring, of the image, which has important implications for the spatial frequency content of the image (Schowengerdt, Park and Gray, 1984). In particular, nearest neighbor resampling is equivalent to convolution with a rectangle function and bilinear interpolation is equivalent to convolution with a triangle function (Shlien, 1979; Billingsley, 1983). Both suppress the higher spatial frequencies leading to loss of resolution in the reconstructed image. The parametric cubic convolution interpolator preserves more of the higher frequencies but can exhibit ringing near edges. Cubically convolved imagery still has a somewhat blurred appearance compared to unresampled imagery.

A resampling method called "restoration" or "deconvolution." attempts to recover losses suffered in the image due to the imaging process itself (Dye, 1975; Wood, Schowengerdt and Meyer, 1986; Wood, 1986). The grey level value assigned to a particular pixel in an image is the result of averaging or integrating information from a neighborhood of ground radiance values (Kalman, 1984). This blurring effect is inherent to all imaging systems and can be described in terms of the system's point spread function, or impulse response function, which characterizes the irradiance distribution at the image plane of an object which is an ideal point source. A detailed knowledge of the system's point spread function can be used as a model from which to make linear combinations of pixels and their neighbors so that a new point spread function can in effect be synthesized which can approach a resolution limited by the original sampling interval. Restoration is different from all the other resampling methods discussed here, which are attempts to reconstruct the original image before sampling, rather than the scene from which the image was derived.

Using restoration for resampling would boost, rather than suppress, those frequencies which had already undergone some suppression in the imaging process. The idea is to use a restoration filter as the interpolator to account for blurring that had already occurred, rather than to introduce more blurring (high frequency suppression) in the image. The approach is designed to predict the radiance value that would have been obtained using a sensor with a more desirable point spread function positioned directly over the location for which a value is desired (Dye, 1976). The result of restoration resampling is a sharper image with greater information content (Kalman, 1984). Restoration has been shown to give better classification results than other resampling methods under certain combinations of blur and noise (Lai, et al., 1984). Resolution can be improved in along-scan Landsat Multi-Spectral Scanner (MSS) data from an effective instantaneous field of view of 86 meters to one of 58 meters (Schowengerdt and Wood, 1986). (It is important to note, however, that the effective instantaneous field of view is only one of many measures of resolution and does not account for such degrading artifacts as noise enhancement and edge grey level overshoot. Also, the presence of the non-linear Butterworth filter in the sensor's along-scan direction makes for more potential resolution improvement than could be expected along-track where the filter is not used.) Restoration requires about the same amount of computer time as cubic convolution.

Restoration must not be confused with edge enhancement which is basically a heuristic procedure designed to enhance visual discrimination of features. Restoration also enhances edges, but it does so in a physically meaningful way which results in less radiometric error. Restoration enhances edges by removing known blur while minimizing radiometric error.

Restoration is possible because typical satellite sensors have sampling rates (the number of samples per instantaneous field of view) greater than one. For the MSS there are 1.31 samples of the 63-by-63 meter ground projected instantaneous field of view along-scan and 0.93 samples along track. However, the imaging process itself degrades the image further. For instance, although the ground projected instantaneous field of view for MSS is 63-by-63 meters based just on the scanning aperture, the image-forming optics and electronic filter increase it to 77-by-65 meters (the filter is effective only along-scan). Sample scene phasing increases it to 86-by-122 meters and bilinear resampling increases it even further to 104-by-148 meters (Park et al., 1984). These combined effects conspire to produce redundant information from pixel to pixel which can be exploited using restoration techniques, since the blur added by these components effectively increases the sampling rate.

Degradations accounted for in modeling sensor systems for restoration usually include sensor parameters such as those describing the optics and electronic filters, but may also include motion blur, atmospheric effects and ground processing. Thus, unlike the other resampling methods, restoration is sensor dependent and possibly even acquisition dependent. However, since no real physical system can be inverted exactly, noise enhancement can occur when the restoration is pushed toward its theoretical limit. The major drawback for restoration, though, is that it requires a thorough understanding of the sensor design and a mathematical model of the sensor's point spread function based on detailed engineering specifications. Thus it requires a major development commitment. The model only needs to be developed once for each sensor though (unless atmospheric or other time varying effects are to be accounted for) and once made the associated resampling coefficients can be used on all images derived from that sensor.

The restoration used at ERIM is the linear least squares deconvolution approach (Dye, 1976). Restoration (or deconvolution) is performed as a linear filter acting on the original data. The coefficients are chosen so that the difference between grey level values produced using the correctly located desired point spread function and those produced using the linear filter approximation are minimized in a least-squares sense. This process has been shown to produce image data with higher radiometric fidelity than can be achieved with cubic convolution (Shah and Wilson, 1977).

5.1.2 Radiometric Balancing

In order to make meaningful comparisons between the images from different sensors, careful radiometric processing of imagery needs to be made. The goal is to create an image from one sensor which can be used interchangeably with an image from another sensor without changing the interpretation methods.

The need for radiometric balancing between imagery from different sensors may arise if over the long term users must deal with radiometric changes in a sensor due to sensor aging. Eventually, when a sensor becomes inoperative and a new sensor with different spectral characteristics takes its place, the ability to correct the data from the new sensor to match the old (or visa versa) would become important.

Good methods for radiometric balancing between sensors would also be applicable to balancing between multi-temporal same-sensor data. This is necessary in change detection and for producing large area mosaics. In both cases, radiometric matching problems arise because the images are acquired at different times and atmospheric conditions or the sun angle may have changed.

Several approaches for multitemporal/multisensor radiometric balancing have been either proposed or successfully implemented at ERIM. They are discussed below.

5.1.2.1 Band Emulation

One spectral band may be emulated by a linear combination of the other bands; the coefficients may be found through the principal components or tassled cap transforms and a multivariate regression relation (Suits, et al. 1988). Thus, a band from one sensor can be emulated from spectrally similar bands of another sensor. Alternatively, a complete multivariate regression procedure could be used, thereby avoiding the transformations altogether; ancillary data (e.g. sun angle, weather, haze) could be used to increase the accuracy of the emulated band (e.g. Odenweller and Rice, 1988). Using this approach, quantities such as vegetation index and albedo would be a function of the actual reflectance characteristics of the objects in the scene and the spectral band emulation quality. The multivariate approach was used successfully at ERIM to emulate Landsat Multi-Spectral Scanner (MSS) false color images from four other satellite sensors, Landsat Thematic Mapper (TM), NOAA's advanced very high-resolution radiometer (AVHRR), the coastal zone color scanner (CZCS) and the SPOT high resolution visible (HRV) sensor (Suits, et al. 1988). However, the use of simulated signals does raise the possibility of creating signals which do not represent the physical world. Procedures must be devised for testing the emulated signals for realism.

5.1.2.2 Physical Model

The radiometric values in an image are a function of the target reflectances and illuminating conditions. It should be theoretically possible to transform the data (radiance) values into reflectance values. Several factors would have to be taken into account, including target reflectance, sun elevation angle, view angle, atmospheric transmittance and radiance (amount and type of suspended particles, water vapor concentration, etc.), between-sensor calibration and absolute calibration (Malila and Crist, 1985). A physical model which takes these parameters into account could be used to reduce all the data values in an image to common reflectance values. One such model has been proposed (Shah, 1988) but never implemented. It may be that many of these effects are currently beyond our ability to correct, or that sufficient ancillary information would not available to make an adequate correction anyway.

5.1.2.3 Categorization Approach

The imagery could be categorized based on the bands actually present in the sensor. Then, when the ground cover spectral signatures, sun angle and other relevant parameters are known, the correct reflectance could be assigned to each pixel. The drawback of this method is that each image would have a blocky appearance as a result of the categorization. One way to deal with this would be to artificially add texture to the image, again based on the known land cover types. The feasibility of this texture approach had never been investigated at ERIM.

5.1.3 Combining the Data

The most common method for combining registered and radiometrically normalized data from different sensors is sometimes call "band sharpening." The idea is to use the data from one sensor with good spectral resolution along with that of a sensor with good spatial resolution in such a way as to exploit the best features of each.

The simplest way to do this is by band replacement. Figure 1 shows a combination of a Landsat-4 Thematic Mapper (TM) image and a STAR-1 SAR image. Both images have been geometrically corrected and resampled to a common ground plane. The STAR-1 data (X-band) and TM data (bands 1 through 4 at 0.45-0.52, 0.52-0.60, 0.63-0.69, and 0.76-0.90 μm) werefilmed and photographically combined to produce both natural color and false color images.

The SAR image provides the user with very fine spatial detail and textural information. The very bright radar returns are reflected from tall buildings, bridges and objects close and parallel to the flight path. The Landsat TM data adds spectral resolution to the radar data. In the lower left of the false color composite TM data there are two quarries that are saturated (white). With the addition of the radar data the interpretability of the area is significantly increased. Another good example is the automobile plant under construction in the upper left of the false color TM image. This image enhancement technique is also useful for trafficability analysis (Fox, 1988).

Another method for band sharpening is to first transform the multispectral data from Cartesian red-green-blue (RGB) into another space, make the replacement there, then inverse transform back to red-green-blue space. The most commonly used transform is hue-intensity-saturation (HIS) or hue-saturation-value (HSV). The intensity or value is replaced by the higher resolution image and the result inverse transformed.

In the HSV transformation, the data in Cartesian RGB coordinates is transformed into cylindrical HSV coordinates, such that the line R=G=B and the plane defined by H=0 coincide. The transformation is accomplished by performing two rotations that bring the diagonal of the RGB color cube into coincidence with the plane H=0.

This method has been used to sharpen multispectral Landsat data with geometrically registered higher resolution SPOT panchromatic data (Figure 2). The multispectral data is transformed into HSV space and

 p. 48 Blank

SYNTHETIC APERTURE RADAR - MULTISPECTRAL SCANNER DATA FUSION

Detroit, Michigan and Windsor, Ontario

ERIM

STAR-1 SAR
X-Band
4 October 1984

LANDSAT-4 THEMATIC MAPPER
False Color Composite
25 July 1982

THEMATIC MAPPER/STAR-1
Data Fusion
Natural Color Composite

THEMATIC MAPPER/STAR-1
Data Fusion
False Color Composite

These data fusion images consist of STAR-1 SAR data (resolution [illegible] meters) and Landsat-4 Thematic Mapper data (resolution [illegible] meters). Both data sets have been geometrically corrected and resampled to a common ground plane with a resampling interval of 4.2 by 4.2 meters. The STAR-1 data (X-Band) and TM data (bands 1-4, 0.45-0.52, 0.52-0.60, 0.63-0.69, and 0.76-0.90 [illegible]) were filtered and [illegible] combined to produce both natural color (TM bands 1, 2, and 3) and false color (TM bands 1, 2, 3, and 4) data fusion images.

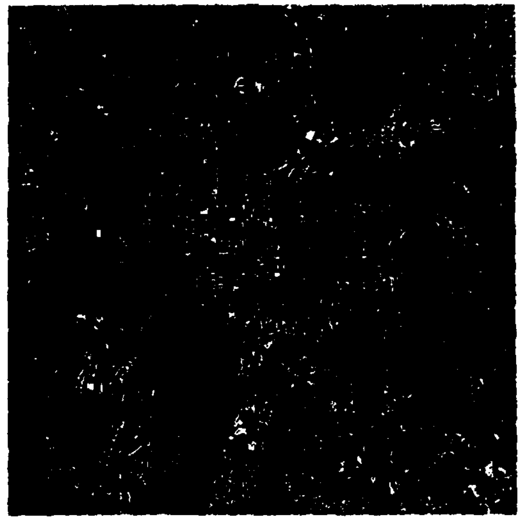
p. 50 Blank

Figure 1

SPOT DATA FUSION

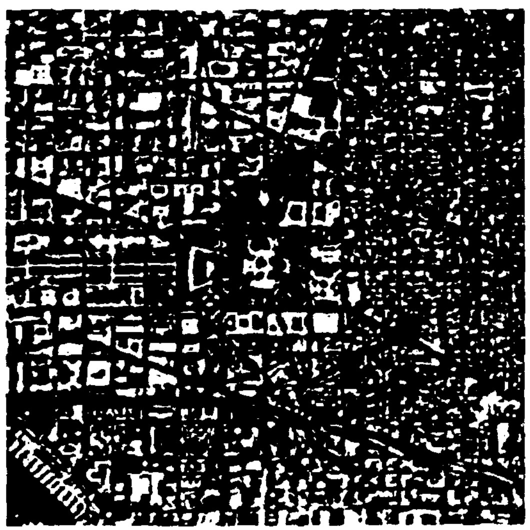
WASHINGTON D.C.

1:400,000

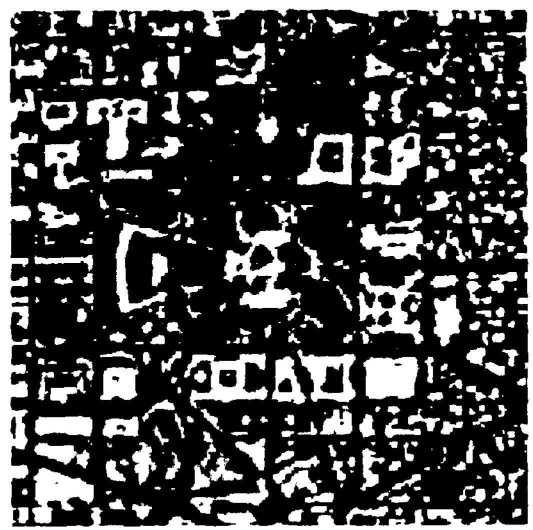


1:200,000

1:100,000



1:50,000



1:25,000

10M PAN MERGED WITH 20M XS
MARCH 1987

Figure 2

p 52 Blank

the value component is substituted with a modified panchromatic intensity. The panchromatic band histogram is stretched so that it is similar to the histogram of the value component of the HSV image. The result is inverse transformed (Nordman, 1988).

5.2 FUSION AT THE FEATURE OR TARGET LEVEL

As previously mentioned, the first issue when developing a multiple sensor surveillance system is to decide where the data association is to take place. Fusing data at the pixel level has already been discussed. Another approach is to store models or templates in symbolic form, convert spatial and/or spectral features to symbolic form, and perform object identification symbolically. A library of feature templates of different types could be stored and the incoming features or feature vectors compared to members of the library.

Declarations from single sensors could be combined by such logical rules a majority voting or weighted summation (Walters, 1989). Thus, each sensor would first process its own data then output its best estimate of target attributes. The ultimate decision would come from combining these "votes." Quantitative measures of evidence could be used for probabilistic (e.g. Bayesian) or possibilistic (e.g. fuzzy set) decisions. The Bayesian approach combines a priori probabilities into an a posteriori probability for decision making. The fuzzy set approach represents not only the value of the evidence but also a measure of the uncertainty associated with it. In the second approach, each sensor would send its raw data to a central processor which would combine it into a master file using all the observation data. A predetermined set of features (or "template) could then be compared to the master file along with some confidence level to make a decision. Even when this approach is used it would be a good idea to keep the raw data as a backup for the photointerpreter to verify the decision.

Once the features (candidate targets) are identified in each sensor's image, a symbolic matching must take place. This could be simply a majority vote or, if some sensors' outputs are more reliable than others', then a weighted vote could occur. If it is determined that a target exists at a certain location, then that location can be highlighted (i.e. cued) in the original imagery for final decision by a photointerpreter. It is generally considered that a human interpreter look at the original or enhanced image, rather than the symbolic extractions, to make the decision, since that is where the context will be found.

5.3 EXAMPLES

5.3.1 Radar and Visible Light Image Data

The combining of pre-processed SAR and multispectral Landsat data has been described in Section 5.1.3. Three-dimensional terrain data can also be extracted from a radar-photo stereo pair. Geometric models of the radar system and camera allow for the reconstruction of projection lines in space which intersect to determine object space positions (Abshier, 1987). For the photograph, the projection lines are rays passing through the photograph at the image points and through the lens at its nodal point. For the radar, the lines of projection are range-Doppler circles formed by intersections of range spheres with Doppler cones. The intersection points determine the three-dimensional object space positions.

5.3.2 Range Data and Reflectance Data

There are several ways of fusing range data and reflectance data. One way is by using a range squared normalization in the reflected energy to negate the decrease in diffuse reflected energy with distance. This has the effect of normalizing collection parameters so that the intensity from continuous objects such as roads is constant.

Laser radar and forward looking infrared, in particular, can be fused in the following way: Small changes within an area measured by laser radar indicate surfaces that could be part of a target. Large changes in range point to possible boundaries. Large infrared changes characterize the relative temperature signature of objects. Thus, one would use small range changes in laser radar data for segmentation and large changes in infrared data for enhancement.

Range is insensitive to the spectral characteristics of the scene as well as to seasonal and diurnal variations. Consequently, the physical drop-offs between the objects and the background create boundaries made up of large range gradients. The gradients inside these enclosed boundaries will be comparatively small. Natural backgrounds such as grass, mud, tree leaves, and sand exhibit random multiplicative noise. The random noise manifests itself in large range jumps. Other natural objects such as trees and shrubs also exhibit large range jumps; the spacing between leaves and branches causes laser radar to record large range differences from one pixel to another. On the other hand, possible targets such as vehicles and buildings present smooth surfaces and slowly varying ranges; therefore, small gradients point to possible targets.

Gradients in the infrared images also enhance the possible targets. Common targets such as tanks and trucks are good infrared sources. Since the objects are the sources, the problems of specular reflections and shadows by illumination processes do not apply (Tong, et al., 1987).

5.3.3 Range Data and Emissive Data

Combining active laser data with high resolution passive forward looking infrared (FLIR) data provides a powerful tool for target discrimination, as this combination gives size and shape simultaneously with thermal contrast and texture (McGlynn, 1989). Essentially, one seeks to exploit the fact that potential targets have a slightly different range of lengths, widths and aspect ratios than most natural clutter. Hence, the first step, after normalizing the range image by computing and subtracting the low frequency trend, is to "prune" the pixel intensity (height) range to the valid range for potential target vehicles, that is, discard pixels outside the valid height range for the targets. Residual high frequency noise can then be attenuated using a median filter (Section 7.4.3.2), and the output thresholded at the expected target height to segment the image into a binary state (Section 3.1). Next, the blobs in the foreground can be reduced to their "skeletal" state by a medial axis transformation. This transformation is used to discriminate the size and shape of each "blob" by successive endpoint reduction of the skeletal state to discard detections which are too small or too large. The resulting detections can then be cued to the original range image for final evaluation by a photointerpreter.

5.3.4 Perspective Views

The addition of terrain information can greatly enhance the interpretability of the imagery.

One way is to generate perspective views through the use of Digital Terrain Elevation Data (DTED) (Schlosser, 1989). This DTED set is a geographically referenced (WGS) terrain elevation matrix (1 degree by 1 degree), with a grid spacing of three arc seconds in latitude and longitude.

DTED files can be used for the generation of perspective views. The DTED is resampled and interpolated where necessary to register it to grey level satellite imagery from, say, SPOT or Landsat. Then the DTED is used to generate a perspective view using the imagery from the sensor with appropriate resolution for the distance from the scene to a preselected viewpoint.

DTED files can also be used for the generation of shaded relief, for slope computations with applications in land use analysis, and for generating colored elevation output where elevation ranges are assigned to specific colors.

Another way of generating perspective views is through the use of two different views of the same scene. ERIM has developed both perspective view generation software (Dye, 1989) and flight trajectory software (Leadholm, 1989). Combined, this software allows a user to define points (x,y,z) on a digitizing table along with an aircraft flight path, as well as viewing locations that a pilot in the aircraft would look toward when the aircraft arrived at those pre-selected points. The perspective view is created, as are the views that the pilot would see as he flew smoothly through the points along the chosen trajectory and his vision panned from the aircraft trajectory to the pre-selected "look" points.

ERIM has also developed a file structure to provide storage for multiple resolution, multi-sensor geographically distributed information over large areas of the globe, called the "Global Data Base" (Dye, 1988a). The perspective view software can be linked to a global data base at the level appropriate for the sensor resolution and viewing distance. A typical perspective view may link 35 or more global data base files having various resolution levels. As the distance from the viewpoint to the terrain increases by a factor of two, the data base files used in the view automatically shift to a lower level (resolution). Perspective views are computed along the flight trajectory at a rate of about seven to 15 minutes per view on a Vax 11/780 computer. Complex objects can be added using Digital Feature Analysis Data (DFAD) by adding solid models using volume elements (or voxels) riding on the terrain surface features (Dye, 1988b).

ERIM has been using the Global Data Base to generate simulated flights within the Las Vegas, Nevada region. This data base is composed of the continental boundaries contained in the World Bank II at 40 kilometer resolution, DTED resampled at 80 meter resolution, Landsat MSS at 50 meter, SPOT at 10 meter and NHAP aerial photographs at 5 meter resolution. This data base permits views to be generated along a trajectory originating in outer space and terminating on earth targets.

Although it is common to generate perspective views from two acquisitions from the same sensor, it is not necessary that that be the case, as shown in the proprietary study done by Abshier (1987). In this work, it was shown that an image can be generated from radar data and a photograph, even if the two sensors were on the same platform.

p 58 Blank

6.0 TARGET AREAS CUEING

After feature detection and extraction and target discrimination, the candidate targets can be cued to the photointerpreter for a final targeting decision. Typically the cued area is highlighted in the raw or pre-processed image (Walters, 1989). This can be accomplished using a box or polygon, a cursor, or by highlighting the candidate targets. In any case, the photointerpreter should be able to turn off or blink the cue so that it does not obscure the view of the candidate target.

Coarse resolution sensor data can be used to detect potential target areas to search in fine resolution imaging sensor data. For instance, FLIR imagery is hard to use at the discrimination level because FLIR images of targets are shadowless and have degrees of uncontrollability for time of day, recent history of the target and so forth. Such irrelevant information and variability makes FLIR imagery difficult to use for target discrimination. However, FLIR imagery can be used to locate small areas of interest (high false alarm and high detection probability, yet small total area compared to the region searched). These cued areas can then be searched with, say, a 3-D laser sensor using only short and highly constrained bursts of active radiation (Brown and Swonger, 1988). Shape is a high quality discriminant when there are many resolution cells (pixels) over the target.

p. 60 Blank

7.0 DISPLAY CONSIDERATIONS

In this section, a broad view of display considerations will be taken to include any method or process which could be applied to ease the interpreter's burden before coming to a decision about the contents of an image.

7.1 DISPLAY DEVICE AND AMBIENT LIGHTING

The display device itself should be chosen to give photointerpreters maximum ease of interpretation of the imagery as well as efficient integration of the display within the confines of the operational area (Frizzell, 1989). Effective display of SAR imagery, for instance, requires that the display monitor be capable of a wide range of output luminance, corresponding to the need to use the display in a variety of ambient lighting conditions. Increases in intensity produce a larger spread of the spot size in some displays (such as color, or multiple gun displays) compared to others (such as monochrome or single gun displays). This spread of spot size acts to lower the resolution. Also, if it is expected that images will be interpreted under conditions of high ambient illumination, some monitors may be inadequate. Finally, high resolution (1024-by-1024) displays, while giving the photointerpreter more context and better resolution, can also result in a lack of focus since there is so much more information to evaluate. The display device, then, needs to be chosen with careful attention to the application and to the environment in which it is to be used.

7.2 DISPLAY METHODS

The display should have the capability to support split screens. (If this is not possible then multiple displays can serve the same purpose). One use for split screens would be to have the target areas from each sensor displayed simultaneously on one display. Another use would be as aid in setting thresholds, for instance when using the SIFT filter (Section 7.4.3.3) or implementing a level slice (Section 3.1). Also, multispectral imagery could be analyzed jointly in the spatial and spectral domains (Rice, 1987) for rapid classification or screening. One window could display the multispectral image while another displayed histograms of each band or scatterplots. The user could designate one or more areas in the image that include the pixels in the class to be studied. Designated pixels could then be used to produce a "highlight" set of scatterplots or would be highlighted in the accompanying histogram. These pixels could be indicated in the image by color coding or surrounding them with a polygon.

Alternatively, the photointerpreter could apply decision rules based on the histograms or scatterplots to determine decision boundaries for filters, classification or thresholding. The user could also apply decision rules in one or more of the scatterplots or histograms in order to help find pixels of interest in the image. The completed selection mask could be used to highlight areas of the image, and the highlighting method should be able to be enabled, disabled or "blinked." The user should be able to modify, add to or delete from the decision rules and the display should respond interactively to these changes. Rules should also be able to be applied to the histograms or scatterplots themselves in order to modify the image, such as changing the gain and bias or applying histogram equalization (Section 7.4.2.2). These histogram modifications ought to be interactive and be able to be applied to each window and each band of each window independently.

Other useful display aids would be the ability to zoom, pan or scroll the image underneath its window. Pixel readout at the cursor position might also be useful for some applications. The ability to overlay images from different sensors in the same window might be useful when the images can be geographically registered. The ability to blink or flicker between images might also be useful in this case.

7.3 INTERPRETATION AIDS

Interpretation aids are aids whose purpose is to help the photointerpreter come to a decision but which do not directly modify the image (as do enhancement aids). For instance, a cursor or template indicating the size and/or shape of the expected target is one such aid. In this case, the cursor itself would have to be tied to the zoom or image resolution so that it remains meaningful across resolution modes. If the expected target is rectangular in shape, the cursor could be orientable or it could be a annulus with the inner radius corresponding to the width of the target and the outer radius to the length. Another interpretation aid would be to have a hard or softcopy set of sample target imagery to help guide the interpreter's decision. The main drawback of such aids is the possibility of the photointerpreter becoming overdependent on the aid. If the target was partially masked by intervening material or camouflage, then these cursor parameters would not be relevant. If the target in the image appeared at a much different orientation than in any of the models, then this might cause an incorrect rejection. This is especially true for SAR imagery in which a small change in orientation of the target with respect to the sensor can change the signature dramatically.

7.4 ENHANCEMENT AIDS

Enhancement aids are a family of processes which modify the displayed imagery. Such aids include zoom, contrast enhancement and filtering.

7.4.1 Zoom

Images can be zoomed using subsampling or other methods to help focus the interpreter's attention. SAR images are typically not subsampled, but rather take either the average of 4 pixels or the maximum.

7.4.2 Contrast Enhancement

Contrast enhancement serves to improve an image based on its contrast and dynamic range characteristics. One application for contrast enhancement is clean-up of an image prior to or after other processing. For instance, the adaptive boxcar filter discussed in Section 7.4.3.4

results in an output image with very little dynamic range. Thus, output from this filter requires a contrast enhancement to display the high frequency content in the image.

7.4.2.1 Gain and Bias

Typically contrast enhancement is performed by applying a single gain and bias to every pixel in the image. The bias is applied as an additive factor, it has the effect of "sliding" the image histogram to the right or left from that of the input image histogram, thus brightening or darkening the image. Care must be taken that the bias is not so large as to saturate at either end, so that pixels "slide" right off one end, unless those pixels are sure not to be pixels of interest. The gain is a multiplicative factor applied to each pixel. It has the effect of controlling the gray level range of the output image histogram. For instance, in a relatively homogeneous region it can be used to increase contrast and dynamic range, thus enhancing the high frequency content of the image. Again, care must be taken to avoid overflow.

Typically gain and bias are implemented either by a switch or an interactive joystick or trackball. In the first case, if it can be determined beforehand that only a few gain and bias adjustments are optimum for the images to be interpreted, then they can be made available on an easy to use switch. If this is not the case, then time must be taken by the photointerpreter to finely adjust the gain and bias using a joystick or trackball.

7.4.2.2 Histogram Equalization

As mentioned above, one of the main drawbacks with gain and bias adjustment is that when attempting to stretch the histogram enough to enhance the detail of interest, many of the pixels can be lost due to saturation. One way around this is to apply a piecewise linear stretch. A different gain and bias are applied to various ranges of grey level values in the image. This often results in a cosmetically more appealing image but at the cost of compromising the radiometric integrity of the image. Also, it may be necessary to apply many different sets of parameters before settling on one that is adequate.

An alternative technique is histogram equalization. In this method, the intensity values in the image are altered such that the resulting image has as nearly as possible a constant intensity histogram. Such images utilize the available display levels well. However, because contrast enhancement is based on the statistics of the entire image, some levels will be used for the depiction of parts of the image which are not of interest, such as the background. Adaptive histogram equalization attempts to overcome these limitations by performing the enhancement at each pixel based on the intensity values of the immediately surrounding pixels. Of course, the drawback here is the huge amount of computation time required. Approximation techniques are typically used instead (Zimmerman, et al., 1988).

7.4.3 Filters

Digital spatial filtering is an important tool both for enhancing the information content of image data and for implementing cosmetic effects which make the imagery more interpretable to the user. Spatial filtering is a context-dependent operation that alters the gray level of a pixel by computing a weighted average formed from the gray level values of other pixels in the immediate vicinity.

7.4.3.1 Adaptive Filters

Traditional spatial filtering involves passing a particular filter or set of filters over an entire image. This assumes that the filter parameter values are appropriate for the entire image, which in turn is based on the assumption that the statistics of the image are constant over the image. However, the statistics of an image may vary widely over the image, requiring an adaptive or "smart" filter whose parameters change as a function of the local statistical properties of the image.

Then a pixel would be averaged only with more typical members of the same population. Adaptive filters, like non-adaptive filters, are used to accomplish various goals including contrast enhancement, edge enhancement, noise suppression and smoothing. Some adaptive filters are described below. An annotated bibliography of adaptive filtering can be found in Mayers and Wood, 1988.

7.4.3.2 Median Filter

The median filter is a popular filter for noise removal. It replaces the center pixel in a neighborhood with the median value of all the pixels in that neighborhood. This filter can also be applied iteratively. The primary benefit of median filtering is that it completely removes small anomalies in an image, even if they are of large intensity, without significantly altering the larger scale features of the image (Kauth and Cicone, 1983). Median filtering is especially useful in delineating the edges in a noisy image in an unbiased way. Of course, the filter window size must be carefully chosen with respect to the target size. Too small a window will result in little filtering and too large a window will result in "blending" the target into the background, thus degrading subsequent detection performance (Miller, 1988a).

The median filter is very effective in removing speckle noise from radar or other single band images. However, it can also remove or obscure small objects of interest. (See Section 7.4.3.3.) For multiple band images the median filter can create non-physically realizable pixels if applied one band at a time. This is because the median filter by definition is based on a single-dimensional ordering of the data. Multispectral data is lacking in an *a priori* natural ordering. In order to effectively use the median filter for advanced target recognizers to assist in segmentation of multi-modal imagery (e.g. laser range, reflective and emissive infrared), a working definition of "median" which applies to multispectral data must be made. Essentially, the multidimensional data must be projected onto a single dimension for some applications. The phenomenology of the particular application may define a natural projection which at least partially reduces the dimensionality of the problem. For instance, Landsat 4 data is readily projected onto a two-dimensional subspace by a linear transformation with very little loss of information. Also, the projection may be adaptive; that is, different projections could be used for different segments of the image. This approach has been applied to Landsat data (Horvath, 1986) and the results visually evaluated. The technique suppresses noise while not destroying the spatial and spectral integrity of the data. Single pixel features were removed and large regions were made homogeneous with their edges sharpened.

7.4.3.3 SIFT Filter

The other drawback of the median filter is that is suppresses all small objects in the image. Usually these are noise, but sometimes they are not. The SIFT (Selectable Iterative Flexible Topology) filter was designed specifically for noise attenuation and speckle reduction in radar imagery but it does not effect pixels that are likely to be of interest to the photointerpreter (Lake, 1989). The filter requires a pre-determined threshold. For this reason alone it is listed under the section "Display Considerations" rather than "Pre-processing" (Section 2.0) because generally a human will be required to set the threshold. If enough is known about the target ahead of time to set a fixed threshold, then this method could be done in pre-processing.

The SIFT filter replaces the center pixel value by the center plus slope if there are at least a specified number of neighbors with grey levels greater than center plus slope but less than center plus slope plus threshold. The effect of setting the threshold insures that pixels above it will not be effected by the filter. It suppresses the noise as well as the median filter, but does not suppress pixels of interest as determined by the threshold. It is a computationally intensive filter and requires a Cytocomputer or other parallel processor implementation.

7.4.3.4 Adaptive Boxcar Filter

The boxcar filter was developed to enhance contrast in basically homogeneous regions such as sand and water. Traditionally, this involved passing a rectangular kernel over an image, calculating the mean of the pixel values inside the kernel (image smoothing), calculating the difference between the average and the pixel value in the center of the kernel (high pass filtering), and adding a percentage of this difference back to the pixel value in the center of the kernel (edge enhancement) (McDonnell, 1981; Eliason and Soderblom, 1977). The filter attenuates the low-frequency, large-scale, shading that dominates the dynamic range of the available grey levels and enhances the high-frequency content. However, the filter also introduces a "ringing" or "haloeing" artifact, particularly evident at boundaries between distinctly different land features.

An adaptive high-pass filter whose parameters change as a function of the local statistical properties of the image is desirable. In particular, a pixel would be processed only with more typical members of the same population. In 1987, a cooperative research effort between the EROS Data Center and the Environmental Research Institute of Michigan (ERIM) was undertaken for developing an adaptive filter for image enhancement that would incorporate local image statistics. The filter has the effect of enhancing detail in homogeneous regions while avoiding the ringing artifact. This filter is discussed in detail in Torres et al., 1988 and Mayers, et al., 1988.

8.0 BIBLIOGRAPHY

Abshier, J., "Radar/Photo Stereo," ERIM Proprietary Internal Memo #OSL-87-4770, 18 September 1987.

Becher, W. D., "Cytocomputer^{TM}, A General Purpose Image Processor," ASEE 1982 North Central Section Conference, April 1982.

Billingsley, F. C., "Data Preprocessing and Processing," in Manual of Remote Sensing, 2nd Ed., Chapter 16, American Society of Photogrammetry (1983).

Brown, W. M. and C. W. Swonger, "A Prospectus for Automatic Target Recognition," ERIM Report - Document Center Number 830080, June 1988

Crane, R. B., W. Richardson, R. H. Hieber and W. A. Malila, "A Study of Techniques for Processing Multispectral Scanner Data," ERIM Technical Report 31650-155-T, September 1973.

Crimmins, T. "Speckle Reduction Algorithm," ERIM Internal Memo SED-82-144-156840, 24 August 1982.

Crimmins, T., "Geometric Filter for Speckle Reduction," Applied Optics, 24, pp. 1438-1443, 15 May 1985a.

Crimmins, T., "Geometric Filter for Reducing Speckle," Proceedings of the SPIE: International Conference on Speckle, 556, pp. 213-222, August 1985b.

Crimmins, T., "Geometric Filter for Reducing Speckle," Optical Engineering, 25, pp. 651-654, May 1986.

Crimmins, T. "Geometric Speckle Reduction" ERIM Final Report 178100-6-F, October 1986b.

Crist, E. P. and R. C. Cicone, "A Physically-based Transformation of Thematic Mapper Data - The TM Tassled Cap," IEEE Transactions on Geoscience and Remote Sensing, 22, pp. 256-263, May 1984.

Dye, R. H., "Restoration of Landsat Images by Discrete Two-Dimensional Deconvolution," Proceedings of the 10th International Symposium on Remote Sensing of Environment, pp. 725-729, October 1975.

Dye, R. H., "Design for a Global Data Base File Structure for Geographic Raster Information," ERIM Internal Research and Development Final Report, February 1988a.

Dye, R. H., "Adding Complex Objects to the Global Data Base," ERIM Internal Memo #IPTL88-RD-04, 29 July 1988b.

Dye, R. H., Private communication regarding software developed under Internal Research and Development funds, 9 May 1989.

Eliason, E. M. and L. A. Soderblom, "An Array Processing System for Lunar Geochemical and Geophysical Data," Proceedings of the Eighth Lunar Science Conference, 1977, pp. 1163-1170.

ERIM Interim Technical Report, 177200-21-T, "Cruise Missile Advanced Guidance (CMAG) Program: First Interim Report," February 1986 performed for the Department of the Air Force, Air Force Systems Command: Aeronautical Systems Division, Wright-Patterson Air Force Base, Dayton, Ohio (no author cited).

Fox, W. C., "Data Fusion," ERIM Internal Memo #IAEL88-CWF-1, 31 March 1988.

Frizzell, William, "Color/Monochrome Display Evaluation," Proposed test plan for classified photointerpreter evaluation, 5 February 1989.

Gleason, J., "Phase III: Radar Image Understanding," ERIM Final Report Document Center Number 8730119, January 1988.

Holmes, Q. A., Private Communication, 27 March 1989.

Holmes, Q. A. and R. E. Sampson, "Real-Time Image Processing using Mathematical Morphology," Manuscript submitted to an upcoming special issue of Photogrammetric Engineering and Remote Sensing on image processing to appear in September 1989.

Horvath, R., "Final Report on IR&D Project 659119 - Multispectral Median Filter," ERIM Internal Memo #IPD-RH-86-2653, 19 February 1986.

Horwitz, H. M., P. D Hyde, and W. Richardson, "Improvements in Estimating Proportions of Objects from Multispectral Data," ERIM Technical Report 190100-25-T, prepared for NASA, April 1974.

Horwitz, H. M., J. T. Lewis, and A. P. Pentland, "Estimating Proportions of Objects from Multispectral Scanner Data," ERIM Final Report 109600-13-F prepared for NASA, May 1975.

Kalman, L. S., "Comparison of Cubic Convolution Interpolation and Least-squares Restoration for Resampling Landsat MSS Imagery" ASP Annual Meeting 51st, 2, March 1985, Falls Church, VA, pp. 546-556.

Kauth, R. and R. Cicone, "Multispectral Median Filter, IR&D Program," ERIM Internal Memo #IPD-RK-83-1653, 21 July 1983.

Kryskowski, D., Private Communication, 21 June 1989.

Kuschel, S. A., "SAR Imagery Recognition Algorithms," ERIM Memo, 4 March 1988.

Lai, D. C., J. Potenza, and K. Verfaille, "Evaluation of Image Restoration Filters for Machine Classification," Optical Engineering, 23, pp. 794-800, 1984.

Lake, C., Private communication and demonstration (The work is undocumented except for "help" files in the Cytocomputer implementation), 5 May 1988.

Leadholm, T., Private communication, 15 May 1989.

Malila, W. A., R. B. Crane, C. A. Omarzu, and R. E. Turner, "Studies of Spectral Discrimination," ERIM Technical Report WRL 31650-22-T, prepared for NASA, May 1971.

Malila, W. A., R. B. Crane, and W. Richardson, "Discrimination Techniques Employing Both Reflective and Thermal Multispectral Signals," ERIM Technical Report 31650-75-T, prepared for NASA, November 1973.

Malila, W. A., R. B. Crane, and R. E. Turner, "Information Extraction Techniques for Multispectral Scanner Data," ERIM Technical Report WRL-31650-74-T, prepared for NASA, June 1972.

Malila, W. A. and E. P. Crist, "Spectroradiometric Transforms and Data Compression," Proceedings of the 19th International Symposium of Remote Sensing of Environment, 1985, pp. 87-94.

Malila, W. A., R. H. Hieber, and R. C. Cicone, "Studies of Recognition with Multitemportal Remote Sensor Data," ERIM Final Report 109600-19-F prepared for NASA, May 1975.

Mayers, M. and L. Wood, "Selected Annotated Bibliographies for Adaptive Filtering of Digital Image Data," USGS Open File Report 88-104, 1988.

Mayers, M., L. Wood, and J. Hood, "Adaptive Spatial Filtering," Proceedings of the ASPRS Symposium, September 1988.

McDonnell, M. J., "Box-filtering Techniques," Computer Graphics and Image Processing, 17, 1981, pp. 65-70.

McGlynn, J., Private communication describing work done at ERIM for a defense contractor (Documentation for this work is classified), 10 May 1989.

Morita, Y. et al., "Identification and Screening of Remote Mine Detection Techniques," ERIM Technical Report 138300-22-T, prepared for the U.S. Army Mobility Equipment Research and Development Command, June 1979.

Miller, J. M., "The Likelihood Ratio Detector and Pre-filtering," ERIM Internal Memo SED-88-032, 18 January 1988.

Miller, J. M., "A Statistical Method for Comparing Target Images or Image Chips" ERIM Internal Memo, 22 March 1988b

Miller, J. M., "MTI Streak Detection," ERIM Internal Memo SED-88-203, 20 June 1988c.

Nalepka, R. F. and J. P. Morgenstern, "Signature Extension: An Approach to Operational Multispectral Surveys," ERIM Technical Report 31650-152-T, prepared for NASA March 1973.

Nordman, M. E., "HSV Transformation," ERIM Internal Memo #SSDD88-MEN-3, 12 April 1988,

Nordman, M. E. and L. Wood, Geometric Correction for CD-FLIR and User's Guide, March 1988.

Odenweller, J. B., and D. P. Rice, "Thematic Mapper Data Screening and External Effects Correction," Proceedings of the 21st International Symposium of Remote Sensing of Environment, 1987, pp. 745-754.

Park, S. K. and Schowengerdt, R. A., "Image Reconstruction by Parametric Cubic Convolution," Computer Vision, Graphics, and Image Processing, 23 pp. 258-272, 1983.

Park, S. K., R. Schowengerdt, and M. Kaczynski, "Modulation-transfer-function Analysis for Sampled Image Systems," Applied Optics, 23, pp. 2572-2582, 1984.

Peterson, L. M., D. M. Zuk, R. E. Sampson, and C. C. Aleksoff, "Overview of Active Infrared Imaging Systems," The 6th International Congress on Applications of Lasers and Electro-Optics, November 1988.

Rauchmiller, R., Private communication describing work done at ERIM for a defense contractor (Documentation for this work is classified), 12 May 1989.

Rice, D., "A Needed Image Display/Analysis Capability: IDA Interactive Distribution Analyzer," ERIM Internal Memo #ISD-DR-87-75, 27 February 1987.

Richardson, W. and J. M Gleason, "Multispectral Processing Based on Groups on Resolution Elements," ERIM Final Report 109600-18-F, prepared for NASA, May 1975.

Poller, N., K. Johnson, J. Odenweller, and C. Hay, "Analyst Handbook for the Augmented U.S. Baseline Corn and Soybean Segment Classification Procedure (C/S-1A)," ERIM Procedures Report FC-E1-00723, October 1981.

Schlosser, S., Private communication, 15 February 1989.

Schowengerdt, R. A., Park, S. K., and Gray, R., "Topics in the Two-dimensional Sampling and Reconstruction of Images" International Journal of Remote Sensing, 5, pp. 333-347, 1984.

Schowengerdt, R. A., and L. Wood, "Image Deconvolution/resampling Capability," Final Report to USGS EROS Data Center, under contract number 14-08-001-22361, July 1986.

Shah, N., "Report on Multitemporal/Multisensor Radiometric Balancing Investigation IR&D," Internal ERIM Memo #IAEL88-NS-2, June 1988.

Shah, N. J. and C. L. Wilson, "Multispectral Data Restoration Study," Bendix Aerospace Systems, Internal Report #BSR4246, 1977.

Shlien, S., "Geometric Correction, registration, and resampling of Landsat imagery" Canadian Journal of Remote Sensing,5 pp. 74-89 (1979)

Simon, K. W., "Digital Image Reconstruction and Resampling for Geometric Manipulation," Proceedings of the IEEE Symposium on Machine Processing of Remotely Sensed Data, pp. 3A-1 - 3A-11, June 1975.

Sternberg, S. R., "Greyscale Morphology," Computer Vision, Graphics, and Image Processing, 35, pp. 333-355, 1986.

Suits, G., W. Malila, and T. Weller, "Procedures for Using Signals from One Sensor as Substitutes for Signals of Another," Remote Sensing of Environment, 25, 1988, pp. 395-408.

Thomson, F. and F. Sadowski, "A Study of Atmospheric Effects on Pattern Recognition Devices," ERIM Final Report #193300-62-F, prepared for NASA, July 1975.

Tong, C. W. et al., "Multisensor Data Fusion of Laser Radar and Forward Looking Infrared (FLIR) for Target Segmentation and Enhancement," Proceedings of the SPIE: Infrared Sensors and Sensor Fusion, 782, pp. 10-19, 1987.

Torres, V., L. Wood, and N. Shah, "Adaptive Filtering for Enhancing Geological Features," Proceedings of the 6th Thematic Conference, Houston May 1988.

Walters, J., "AGGREGATE: An Image Region Extraction System", ERIM Internal Memo ETC-87-138, 28 September 1987.

Walters, J., "Multi-sensor/multi-detector Data Fusion," ERIM Memo, 8 May 1989 and private communication, 10 May 1989.

Wessling, T. Private communication describing work done at ERIM for Singer (no documentation exists except that in the computer code) 5 May 1989

Wood, L. Restoration for Sampled Imaging Systems PhD Dissertation, University of Arizona December 1986

Wood, L., Schowengerdt, R. A. and Meyer, D. "Restoration for Sampled Imaging Systems" Proceedings of the SPIE: Applications of Digital Images Processing IX, 697 pp. 333-342 (1986)

Wood, L. and A. N. Sellman "State of the Art Survey of Automated Registration and Control Point Chip Library Systems" ERIM Final Report #323901-1-F August 1987 An ERIM proprietary report prepared for Telespazio, SpA, Rome, Italy

Wood, L., "Correspondence Methods for Image Registration - Review of the Literature," A working draft, May 1989.

Zelnio, E., "Notes on Model-Based Analysis of SAR Imagery," File notes, December 1986.

Zimmerman, et al., "An Evaluation of the Effectiveness of Adaptive Histogram Equalization for Contrast Enhancement," IEEE Transactions on Medical Imaging, 7, pp. 304-312, 1988.